Analysis of an Implementable Application Layer Scheme for Flow Control over Wireless Networks*

Alessandro Abate, Minghua Chen and Shankar Sastry
Department of Electrical Engineering and Computer Sciences
University of California at Berkeley
Berkeley, CA 94720
{aabate,minghua,sastry}@eecs.berkeley.edu

***Abstract*— This paper deals with the problem of congestion control and packet exchange on a wireless network. The mathematical model corresponding to the real protocol is inspired by and extends a known fluid flow scheme for the control of congestion on a wired network. The necessity to introduce a specific wireless model is motivated by the presence of channel error: often this error (due to intrinsic noise or channel corruption) is not known exactly. This motivates the modification of the model by approximating parts of its structure with binary functions, whose switching point is known precisely. These new discontinuous elements, while in practice greatly simplifying the structure of the algorithm (they are endowed with a 1-bit of information), complicate the theoretical analysis of its dynamical properties. We therefore approximate them with continuous functions with limiting convergence: they thus preserve the simple shape and yield themselves to analysis as well. Given this setup, we then investigate the important issues of existence and uniqueness of the equilibrium for the dynamical system, and of local asymptotic stability. Furthermore, we show that this equilibrium solves a concave net utility optimization problem, of which the classical one for wired networks is a special case. The take away point of this work is that the scheme we propose to handle the traffic on a wireless network is not only innovative and meaningful, but has also the potential to be modified and translated into practical implementation.**

I. Introduction

Transmission Control Protocol (TCP) has recently been the focus of much research (originated, among the many contributions, in [1], [2], [3]). Not long ago, this practical scheme has been dynamically modeled via a system of continuous time differential equations that describe the evolution of the rates (that is, the number of bits per second) of a set of users that exchange information over a network. This is an instance of fluid flow model (see [4], [5]). The study of this model further understanding of the intrinsic characteristics and dynamical properties of the system. Investigating this scheme has nevertheless proven to be a rather challenging task, mostly because of the presence of strong non linearities in the functions that come into play, and because of the distributed nature of the scheme. Moreover, the multiple couplings between its entities (senders, receivers and links) hampers the global understanding of its behavior.

The current fluid flow models for TCP have only been dealing with the case of wired networks, [4] [5]. Fundamental properties such as uniqueness of the equilibria and stability have been studied, [6] [7], and conditions for achieving robustness to disturbances, [8], and to delays, [9], have been introduced.

Quite recently some researchers have turned their attention to the wireless scenario. This new setting poses new, unconfronted challenges, due to the presence of intrinsic noise and channel errors at the link level. An algorithm known as MULTFRC (see [10]) and proposed for video streaming over wireless networks, has introduced a scheme to be applied to TCP-friendly rate control (TFRC) for wireless networks. In [11], a corresponding continuous-time model is introduced and studied. Many properties, such as global stability, robustness conditions to delays and to disturbances, have been derived, [12] [13].

This paper takes a step forward: the presence of channel error is the cause of imperfect feedback from the network to the users; these errors prevent the exact measurement of the congestion status on the network. This motivates the introduction of a simplifying approximation for that part of the model which is affected by noise. This approximation, in the form of a step function that switches at a known (or computable) point, is on the one hand simpler, but on the other hand discontinuous. Because of this, it is quite hard to do analysis on the modified scheme. Some continuous approximations are then introduced, and their limiting behavior studied. With these modifications, the new scheme is prone to yield interesting results.

The paper unfolds as follows: after a brief explanation of fluid-flow models for wireline networks and a concise introduction to the TCP scheme for wireless networks, we propose its related modification and the corresponding continuous approximations. A series of facts will elucidate the existence and uniqueness of the equilibrium for the approximation of the modified system. Furthermore, local stability for the scheme will be proved, and limiting behaviors explained. It will then be shown that the equilibria of the modified model are the solution of a concave net utility optimization problem, of which the generic one proposed by Kelly for TCP on wired netowrks, [4], is a special case. The implications of these results will follow, and a description of future work will close up the paper.

*This work was supported by NSF grant ANI-9905799, the AFOSR contract F49620-00-1-0327 and the NSF grant CCR-0225610.

II. A Practical Flow Control Scheme

In this section we first introduce the dynamical model of the well-known general flow control problem first introduced by Kelly et. al., [4]. Starting from the wired scenario, we motivate and build up the extensions for the more challenging wireless case; finally, a modification to this last model is discussed in order to simplify it and enable its practical implementation.

A. Wired Networks

A communication network is described via a set J of *links* and a set R of *users* (sender-receiver pairs). Each $j \in J$ has a finite capacity $C_j < \infty$. The network interconnections are described via a routing matrix $A = (a_{jr}, j \in J, r \in R)$, where $a_{jr} = 1$ if $j \in r$. A fluid-flow, continuous-time model for the TCP scheme (see [4]) has been proposed in order to facilitate the analysis of the properties of the protocol. To each user a sending rate $x_r \geq 0$ and a utility function $U_r(x_r)$ are associated. $U_r(x_r)$ is assumed to be increasing, strictly concave and $\mathcal{C}^1$. The exchange of information between users over the links can be interpreted as a concave maximization problem (see, among the others, [14] [15]), dependent on the aggregate utility functions for the rates and on some costs on the links:

$$\max \sum_{r \in R} U_r(x_r) - \sum_{j \in J} P_j \left(\sum_{s: j \in s} x_s \right), \qquad (1)$$

where the cost functions $P_j(\cdot)$ are defined as

$$P_j(y) = \int_0^y p_j(z)\, dz. \qquad (2)$$

The terms $p_j(y)$ can be interpreted as "prices" at the link and are assumed to be non-negative, continuous and increasing functions; they represent some congestion measure and, as can be inferred from their structure, they have a local dependence on the aggregate rate passing through the link. As in [5], in this paper we shall stick to the following "packet loss rate":

$$p_j(y) = \frac{(y - C_j)^+}{y}. \qquad (3)$$

Flow control can then be regarded as a dynamical systems problem, dynamically evolving according to the problem 1, that is having an equilibrium which is the solution of 1. User r will accrue a packet loss rate which, under our assumptions of small p_j, can be approximated as $\sum_{j \in r} p_j(\sum_{s: j \in s} x_s)$. The rate control scheme has the following shape:

$$\frac{d}{dt} x_r(t) = k_r \left(w_r^o - x_r(t) \sum_{j \in r} p_j \left(\sum_{s: j \in s} x_s(t) \right) \right), r \in R \qquad (4)$$

with k_r a positive scale factor affecting the adaptation rate, and the constant w_r^o can be physically interpreted as the number of connections that the user establishes with the network; as discussed, the congestion signal (packet loss rate) depends on the sum of the prices along all the links that are crossed by the user. Interpreting the model (4) as a dynamical relation, it is easy to express its equilibrium in an implicit form. In [4] it is shown by Lyapunov arguments that this equilibrium is unique and asymptotically stable. Moreover, the schemes can be endowed, under conditions over their parameters, with many interesting properties (as an example, robustness).

B. Wireless Networks

Wireless channels are affected by errors, due to the corruptibility of the signals flowing through them and to the presence of noise. This directly influences the packet loss at each link in a TCP-like setting. We thus encompass this fact within a new price function for, say, link j:

$$q_j \left(\sum_{s: j \in s} x_s(t) \right) \triangleq p_j \left(\sum_{s: j \in s} x_s(t) \right) + \epsilon_j \geq p_j \left(\sum_{s: j \in s} x_s(t) \right). \qquad (5)$$

This function accounts for both the congestion measure (dependence on the term p_j) as well as the channel error ϵ_j. The TCP model (4) will then depend on this new function $q_j(y)$. It is again easy to calculate the equilibrium of this new dynamic relation; the dependence on the new price function will give a result that is different than the one derived from (4). Interpreting this fact through an underlying optimization problem, as in (1), shows that the new equilibrium will be suboptimal. This fact motivates the introduction of an enhancement to the wireless scheme, as described in the following section.

C. A new Control Scheme for Wireless Networks

In [12] [13] [11], we introduced two extensions to the TCP scheme, both aimed at compensating the suboptimality of the equilibrium point of the rates of the interconnection (4-5) with respect to the simpler scheme (4). In this paper we shall focus on one of these two proposed schemes, the "dynamic update." Assume the term ω_r is time dependent, $\omega_r(t)$, and evolves according to:

$$\frac{d}{dt} w_r(t) = c_r \left(w_r^o - w_r(t) \frac{\sum_{j \in r} p_j(\sum_{s: j \in s} x_s(t))}{\sum_{j \in r} q_j(\sum_{s: j \in s} x_s(t))} \right). \qquad (6)$$

We can interpret this dynamical relation, in a fluid-flow sense, as the modification of the number of connections that the user has with the network. It is easy to compute the equilibria of this new interconnection (the couple $\{x_r, \omega_r\}$), and check that the "optimum point" of the rates is the same as that of (4). Aiming at a dynamical analysis of this scheme, in [11] we showed that the interconnection is globally asymptotically stable, under the realisitc assumption that the two dynamical relations evolve in two different time scales. In [12] and [13] we instead investigate the robustness of the scheme to delays and study its resilience against disturbances. It is important to notice that this scheme can be easily implemented by adjusting the number of connections which an application opens in a real network. Therefore, it is an *application layer based approach* and it is easy to deploy, as it does not require changes on the network's infrastructure or its protocol.

D. The Indicator Function

From the structure of Eqn. (6) we can gather that the implementation of the control law on w_r depends on the precise measurement of the ratio $\frac{\sum_{j\in r} p_j(\sum_{s:j\in s} x_s(t))}{\sum_{j\in r} q_j(\sum_{s:j\in s} x_s(t))}$, which is the portion of the end-to-end packet loss rate that is exclusively caused by congestion.

From an end-to-end point of view, users can infer which packet is lost only by observing a discontinuity in the sequence number that is carried by every packet[1]; the reason of the loss (congestion or channel error) would not be given though. Therefore users can only precisely measure the end-to-end packet loss rate, i.e. $\sum_{j\in r} q_j(\sum_{s:j\in s} x_s(t))$, but not the quantity due to congestion, i.e. $\sum_{j\in r} p_j(\sum_{s:j\in s} x_s(t))$.

In principle, users can have the ability to exactly measure $\sum_{j\in r} p_j(\sum_{s:j\in s} x_s(t))$, provided more information is gathered from the network infrastructure (like the routes for instance). As an example, the routers and the base stations can generate an Explicit-Loss-Notification (ELN) marking[2] on consecutive packets when they understand that the current packet is lost due to the wireless transmission. Therefore if the users observe a lost packet, they can check the ELN bit on the successive packet to see whether the loss is caused by congestions or by channel error. This way, users can get a precise measure of $\sum_{j\in r} p_j(\sum_{s:j\in s} x_s(t))$, and therefore a better estimate of the above ratio. Other solutions are based on end-to-end statistics, or there exist schemes that are not using packet loss as a congestion measure: for instance, TCP Vegas quantifies the congestion on a measure of the queueing delay. However, to our best knowledge, none of the real world networks infrastructures currently employ these functionalities. Even worst, it is very hard to add this enhancement in every router and base station, and it breaks the end-to-end principle Internet relies on.

All the above motivates the pursue of a better way to control the quantities w_r based on some alternative that is easy for users to measure in reality. We first gauge how the ratio affects the system's performance in (6):

- If a route r is underutilized, then the ratio is zero; this implies that the number of connections $w_r(t)$ increases in order to boost the user rate $x_r(t)$, which makes the system pursue full utilization on the route r;
- If the route r is fully utilized, i.e. if any one of its link is congested, than the ratio takes a value between zero and one, finely adjusting $w_r(t)$, and hence $x_r(t)$, to make the system pursue the maximum utility.

This behavior suggests the idea of replacing the ratio with an *indicator function*. Specifically, let us introduce the following quantity:

$$I_r(x) = \mathrm{Ind}\left(\sum_{j\in r} \frac{(y_j(t)-C_j)^+}{y_j} > 0\right) = \begin{cases} 1, & \text{if route } r \text{ is congested at time } t; \\ 0, & \text{otherwise.} \end{cases} \tag{7}$$

Here $y_j(t) = \sum_{s:j\in s} x_s(t)$ is the aggregate rate flowing through link j. From this definition, we can observe that

- $I_r(x)$ has exactly the same behavior as the ratio when route r is underutilized, therefore, replacing the ratio with $I_r(x)$ will not affect the system's thrust to pursue full utilization.
- If any of the links of route r is congested, $I_r(x)$ does not have the exact same behavior as the ratio; instead, it assumes the value one to push down $w_r(t)$, so as to decrease $x_r(t)$ in order to avoid further congestion on the route.

Unlike the ratio in (6), the value of the indicator function can be easily and accurately estimated by each user. In fact, its value is directly correlated to changes on the round trip time (RTT) for each user[3]. Physically, RTT consists of the round trip propagation delay and round trip queuing delay. For a given route, assuming the backward path is congestion-free, i.e. the incoming rates to the sender are less than the links capacities, the round trip propagation delay is a fixed value, and the queuing delay is zero if the forward path is not congested. If the forward path is congested, the queuing delay increases to positive values, and if the path is continuously congested keeps increasing to a maximum value, i.e. until the buffer is overflown.

Hence, clearly, *an increase in RTT is due to the presence of forward congestion* (and therefore increasing queueing delay)[4]; the increase itself is symptomatic of the indicator function assuming value one. On the other hand, if there is no increase in RTT, then most probably the route is not congested, which means that he indicator function is likely to be equal to zero.

The system (4-5), endowed with this new term, is modified as:

$$\frac{d}{dt}x_r(t) = k_r\left(w_r(t) - x_r(t)\sum_{j\in r}(\epsilon_j + p_j(y_j(t)))\right) \tag{8}$$

$$\frac{d}{dt}w_r(t) = c_r\left(w_r^o - w_r(t)I_r(x)\right). \tag{9}$$

The model (8-9) is a nonlinear, coupled system with discontinuities introduced by the terms $I_r(x), r \in R$. The discontinuities make it difficult to analyze the system, as the new vector field is no longer continuous. As such, it does not fit into the classical framework for analysis previously employed; it would instead require the study of solutions in the *Filippov* sense (see [16]). We decide to tackle this

[1] In practice, the sender waits for three duplicate acknowledgements asking for the retransmission of the missing packet, before it asserts that the packet is lost.

[2] Along with ELN, there exist schemes known as Explicit-Congestion-Notification (ECN) that, as intuitive, work similarly.

[3] Here we call RTT the sum of the time it takes a packet to go from sender to receiver, and back.

[4] Again, under the slack assumption that the incoming rates to the sender are less than the links capacities.

problem by approximating the $I_r(x), r \in R$ with continuous functions; hence we get a continuous approximated version of the system (8-9), which we describe in the following.

III. Continuous Approximations of the System and the Two Times Scale Assumption

The parameter-dependent function we use to approximate $I_r(x)$ in (9) is the following[5]:

$$f_r(x) = \frac{e^{\sum_{j\in r}\ln\left(1+e^{\beta\frac{y_j-C_j}{y_j}}\right)}-1}{1+e^{\sum_{j\in r}\ln\left(1+e^{\beta\frac{y_j-C_j}{y_j}}\right)}},\quad r\in R; \tag{10}$$

furthermore, the discontinuous quantity $p_j(y_j(t)) = (y_j(t) - C_j)^+/y_j(t)$ in (8) is approximated using the following function:

$$g_j(y_j(t)) = \frac{1}{\beta}\ln\left(1+e^{\beta\frac{y_j(t)-C_j}{y_j(t)}}\right),\quad j\in J. \tag{11}$$

It should be clear that $f_r(x) \to I_r(x)$ and $g_j(y_j(t)) \to p_j(y_j(t))$ as $\beta \to \infty$.

The corresponding approximated system is, $\forall r \in R$,

$$\begin{cases} \frac{d}{dt}x_r(t) = k_r\left(w_r(t) - x_r(t)\sum_{j\in r}(\epsilon_j + g_j(y(t)))\right); \\ \frac{d}{dt}w_r(t) = c_r\left(w_r^o - w_r(t)f_r(x)\right). \end{cases} \tag{12}$$

Since the approximated system in (12) is continuous, we can then analyze its equilibrium and stability for arbitrary values of β. As $\beta \to \infty$, the system in (12) approaches the original system (8-9). Therefore, the logic is to analyze the properties of the system (12); moreover, by letting $\beta \to \infty$, we expect to reveal those of the interconnection (8-9).

The approximated system in (12), although continuous, is still hard to analyze in general. Like the model (4-6), it is a nonlinear, coupled, multivariable system, and the two equations are not exactly symmetrical even though they might appear to be so.

In [11] we argue that in the actual TCP schemes the rate of change of the quantity $w_r(t)$, representing the number of connections that a user has with the network, is dimensionally slower than that of $x_r(t)$, representing the source sending rate. Therefore, inspired by control literature on single perturbation systems (for instance, refer to [16]), we carefully make a key assumption to enable the decoupling of the system into two time scales: *the dynamics corresponding to $x_r(t)$ and $w_r(t)$ evolve in two different time scales; the first in a faster one, while the second in a slower one*. This helps us derive strong results on the overall interconnection.

The two time scale assumption applied to the approximated system in (12) highlights two kinds of dynamics: a fast one, which is described in the *boundary-layer system*, and a slow one, which is encompassed in the *reduced-order system*. The fast interconnection is described, $\forall r \in R$, as

$$\begin{cases} \frac{d}{dt}x_r(t) = k_r\left(w_r(t) - x_r(t)\sum_{j\in r}(\epsilon_j + g_j(y(t)))\right), \\ w_r(t) = \text{constant}; \end{cases} \tag{13}$$

in the slower timescale, we instead have the following dynamics, $\forall r \in R$:

$$\begin{cases} x_r(t) = \frac{w_r(t)}{\sum_{j\in r}(\epsilon_j + g_j(y(t)))}, \\ \frac{d}{dt}w_r(t) = c\left(w_r^o - w_r(t)f_r(x(t))\right); \end{cases} \tag{14}$$

Under the two times scale setting, the behavior of the system can be described as follows. On the fast timescale, w_r can be thought as being held constant, and the entire system can be expressed as the boundary system shown in (13). This system is nothing but a slight modification of Kelly's control system on wired network (as expressed in (4)), except for w_r^o replaced by the "constant" $w_r(t)$ and the price function $p_j(y_j(t))$ replaced by $\sum_{j\in r}(\epsilon_j + g_j(y(t)))$; the behavior of the boundary system can thus be easily inferred from the known results of the system in (4). It has a unique and globally exponentially stable equilibrium, which is a function of w_r. Particularly, on the fast timescale, x_r converges to the equilibrium manifold defined as follows:

$$x_r(t) = \frac{w_r(t)}{\sum_{j\in r}(\epsilon_j + g_j(y_j(t)))}.\quad r\in R \tag{15}$$

On the slow timescale, x_r has already converged to the equilibrium manifold, and the system collapses into the reduced system described in (14). Its behavior determines how the approximated system evolves in the long run; therefore, together with the boundary layer system, it fully characterizes behavior of the system for all possible times. Motivated by the above considrations, we shall mainly focus on investigating the reduced system in (14).

IV. Existence, Uniqueness of The Equilibrium, Its Local Stability and The Related Optimization Problem

Given a general nonlinear system, the existence of an equilibrium and its stability represent among the first things to look at.

In this section we show that the system in (12) has a unique equilibrium, and that this equilibrium is locally exponentially stable. We start from showing that any existing equilibrium is locally stable in a neighborhood; then, thanks to this fact and together with some results from the Poincare-Hopf Index Theorem (see [16]), we conclude that there can be only one equilibrium.

Before stating the main results, the following fact is introduced. We simply denote $x(t) = (x_1(t), \ldots, x_{\text{card}(R)}(t))^T$; similarly for $w(t)$.

Lemma 1: The equilibrium manifold shown in (15) is a one-to-one mapping between $x(t)$ and $w(t)$; moreover, the following holds on the manifold:

$$\dot{w} = D(x)\dot{x},$$

[5]For simplicity reasons, we do not make this dependence on the parameter β explicit in the quantities f_r and g_j

where

$$D(x) = \mathrm{diag}(x)\Big(\mathrm{diag}\Big(\sum_{j\in r}\frac{\epsilon_j + g_j(y_j)}{x_r}\Big) + A^T\mathrm{diag}(g_j'(y_j))A\Big)$$

is a product of two positive definite matrices, and as such all its eigenvalues are positive.

Proof: Refer to Appendix A. ■

Remark 1: Lemma 1 implies that within the reduced system (14), analyzing the behavior of the system with respect to x is equivalent to carrying out the analysis with respect to w; as a matter of fact, both of them, as well as their derivatives, are in a one-to-one relationship.

Based on the two times scale decomposition and on singular perturbation theory (see, for instance, [16] or [17]), showing that for the approximated system in (12) any possible equilibrium is locally exponentially stable follows from the argument that both the boundary system and the reduced system need to be locally exponentially stable around the equilibrium. We first claim the following lemma for the reduced system:

Lemma 2: Assume that for the reduced system in (14), x^o is one of its possible equilibria; then x^o is locally exponentially stable for any $\beta > 0$.

Proof: Refer to Appendix B. ■

Furthermore, exploiting the fact that the boundary layer system is locally exponentially stable (see [18]), we apply arguments used in [16] and in [17] for the stability of singular perturbation, non-linear system to infer that any equilibrium of the composite system shown in (12) is locally exponentially stable. This fact is stated in the following:

Theorem 1: If x^o is an equilibrium for the composite system shown in (12) with arbitrary $\beta > 0$, it is locally exponentially stable.

Remark 2: Theorem 1 and Theorem 2 state the existence of a unique equilibrium and ensure its locally stability for the continuous approximated system in (12), for any value of β. At the limit as $\beta \to \infty$, the approximated system approaches the original discontinuous system in (8-9). Therefore, for extremely large β, we expect the approximated system to have a very close behavior to the original system, except right at the discontinuities $y_j(t) = C_j$.

Thus far we have shown that any existing equilibrium is locally exponentially stable. Another important question to address is how many equilibria there are for the system. The answer is stated in the following:

Theorem 2: For any arbitrary $\beta > 0$, the approximated system (12) has one unique equilibrium.

Proof: Refer to Appendix C. ■

In the folowing, we motivate how the unique equilibrium solves a concave optimization problem, which is a modification of the one proposed for the wired case in Eqn. 1.

Theorem 3: For any arbitrary $\beta > 0$, the unique equilibrium of the approximate system in (12), denoted by (x^*, n^*), solves the following concave optimization problem

$$\max_{x\geq 0}\sum_{r\in R} U_r(x_r) - \sum_{j\in J}\int_0^{y_j} g_j(z)\,dz, \qquad (16)$$

with $U_r, r \in R$ being the concave function:

$$U_r(x_r) = \int_0^{x_r} h_r^{-1}\left(\frac{w_r^o}{\nu}\right)d\nu, \quad r \in R,$$

where $h_r^{-1}, r \in R$ is the inverse of the monotonically increasing function h_r:

$$h_r(z) \triangleq \Big(\sum_{j\in r}\epsilon_j + z\Big)f_r(z) = \Big(\sum_{j\in r}\epsilon_j + z\Big)\frac{e^{\beta z}-1}{e^{\beta z}+1}.$$

Proof: First it is easy to see the net utility function in (16) is concave. Then the claim follows by setting to zero the derivative of the net utility function with respective to x. ■

One observation for Theorem 3 is in order: the unique equilibrium for the system in (12) in the wireless scenario solves a concave optimization problem which is similar to the general one (Eqn. 1) solved in the wired network (see [4]), but with different utility functions $U_r(x_r)$ for each user. More prcisely, while the $U_r(x_r)$ in the wired network case is only a function of x_r, in wireless scenario it is also a function of $\sum_{j\in r}\epsilon_j$, that is the packet loss rate associated with the route r. In fact, if we let $\beta \to \infty$ and $\epsilon_j = 0, \forall j \in J$, i.e. if we tend to the wired network scenario, we have $h_r(z) = z$, and thus the optimization problem in (16) becomes identical to the wired network optimization one. In this case, the equilibrium (x^*, n^*) is exactly the same as x^o, implying the optimization problem in the wired network is merely a special case of that in (16).

On an actual implementation of the proposed system (8-9), it is necessary to discretize continuous quantities. For instance, controlling $w_r(t)$ is implemented by adjusting the number of connections, which has to be an integer number; controlling $x_r(t)$ is implemented by adjusting the number of finite packets to be sent out in a time interval. Therefore, it is very unlikely that the system will operate at those points of discontinuity. From this point of view, the analysis based on the approximated system is accurate enough to predict and interpret the performance of the actual implementation of the algorithm.

From a theoretical point of view, the existence of a unique locally stable equilibrium encourages our effort to show that in fact the equilibrium is *globally* asymptotically stable; indeed we have seen that the whole setting can be interpreted as a utility maximization problem that holds globally.

A. Simulations

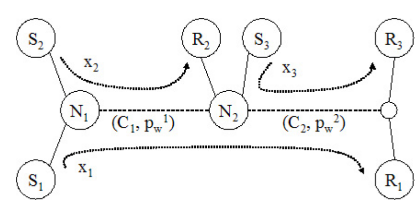


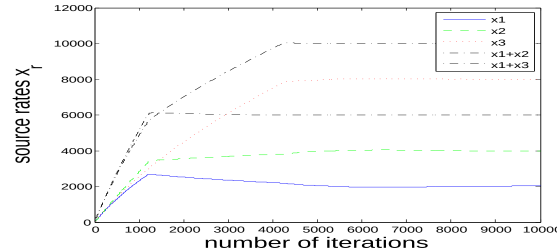
Fig. 1. Simulation topology.

In the following we present the output of some simulations. They show how the performance of the modified scheme closely matches that of the original scheme, in which we assumed full information of the feedback signals from each link. The topology is presented in Fig. 1 and matches that in [11]. The two time scales assumption has been taken into consideration by properly setting the multiplicative constants in the differential equations: it can in fact be observed that the changes of w_r are slower than those of the rates x_r. The initial conditions for Figure 2 are precisely those of Fig. 5 in [11]. The reader should compare these two outcomes to convince himself of the similarity of the results. Due to the discretizations we introduced, the current outcomes display some oscillations (see for instance 2-(b)) that were not present in the original scheme; for this reason, we have refined the integration step and thus necessarily increased the simulation time. This oscillating behavior happens around the optimum for the system (see 2-(a)), which matches that of [11]; we discussed that this optimum corresponds to the full utilization of the links (observe the oscillations of the congestion measures, (see 2-(c))).

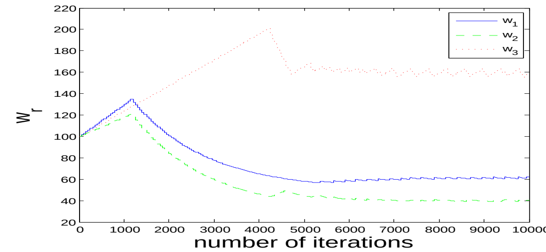
V. Conclusions and Future Work

Standing upon the results presented in [11], where a fluid-flow approximation as a dynamical scheme for controlling the flow over packet-switched wireless networks was proposed and analyzed, this paper introduces an alternative of such model for the wireless scenario. The model is obtained by introducing an indicator function. This simplification is motivated by the necessity to apply the scheme to real world networks, which present inaccurate feedback to the end-users; the new, 1-bit scheme is still an application layer based approach, which therefore does not require any change in the network infrastructure and protocol. The modified model, although easier to implement than its precursor, comes at the cost of introducing some discontinuities in the dynamics, which complicate the theoretical analysis. Therefore, we propose an approximation based on some continuous, parameter-dependent functions which, at the limit, coincide with the discontinuous ones. The new functions yield themselves to some analysis: we prove the existence and uniqueness of the equilibrium of the interconnected systems, solving a concave net utility optimization problem, of which the generic one proposed by Kelly et. al., [4], is a special case. Moreover, we show that this scheme, on a neighborhood of the equilibrium, is exponentially stable. These results are accurate enough to predict and interpret the performance in reality, and are interesting enough to encourage continuing efforts in theoretical aspects.

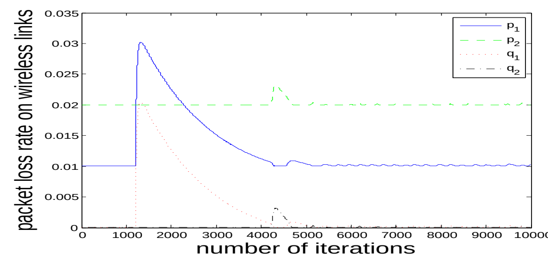
Given the parallel with the model in [11], the investigaton of the global asymptotical stability of the unique equilibrium holds promising results; furthermore, interpreting the properties of the equilibrium from the network optimization standpoint, such as fairness between users and route utilization, may give important insights. The delay stability and the robustness to stochastic disturbance are also interesting and important to investigate from both a practical as well as a theoretical point of view.



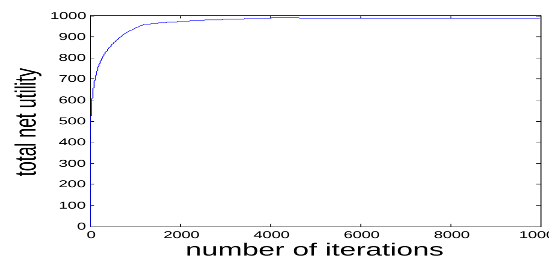
(a)



(b)



(c)



(d)

Fig. 2. Simulations for the modified "dynamic-update" scheme: convergence of (a) rates $x_r(t), r = 1, 2, 3$, (b) $w_r(t), r = 1, 2, 3$, (c) packet loss rate $p_j(\cdot)$ and $q_j(\cdot), j = 1, 2$, and (d) net utility, with initial rate set to 0.

VI. Acknowledgements

The authors would like to thank Ao Tang for the helpful exchange of e-mails, and ChenKe Xu for the discussions.

References

[1] V. Jacobson, "Congestion avoidance and control," *Proceedings of ACM SIGCOMM*, pp. 314–329, 1998.

[2] S. Floyd, "Tcp and explicit congestion notification," *ACM Computer Communication Review*, vol. 24(5), Oct 1994.

[3] S. Floyd and V. Jacobson, "Random early detection gateways for congestion avoidance," *IEEE/ACM Trans. on Networking*, vol. 1(4), pp. 397–413, Aug 1993.

[4] F. P. Kelly, A. Maulloo, and D. Tan, "Rate control for communication networks: shadow prices, proportional fairness, and stability," *Journal of the Operational Research Society*, vol. 49, pp. 237–252, Dec 1999.

[5] F. Kelly, "Fairness and stability of end-to-end congestion control," *European Journal of Control*, vol. 9, pp. 159–176, 2003.
[6] Z. Wang and F. Paganini, "Global stability with time-delay of a primal-dual congestion control," *IEEE Conference on Decision and Control*, Dec 2003.
[7] F. Paganini, Z. Wang, J. C. Doyle, and S. H. Low, "Congestion control for high performance, stability and fairness in general network," *ACM/IEEE Trans. on Networking*, to appear.
[8] R. Johari and D. Tan, "End-to-end congestion control for the internet: delays and stability," *IEEE/ACM Trans. on Networking*, vol. 9, no.6, pp. 818–832, Dec 2001.
[9] G. Vinnicombe, "On the stability of end-to-end congestion control for the internet," *University of Cambridge Technical Report CUED/F-INFENG/TR.398*, 2001.
[10] M. Chen and A. Zakhor, "Rate control for streaming video over wireless," *Proceeding of Infocom 2004*, 2004.
[11] M. Chen, A. Abate, A. Zakhor, and S. Sastry, "Stability and delay considerations for flow control over wireless networks," *UCB EECS ERL Technical Report*, 2005.
[12] M. Chen, A. Abate, and S. Sastry, "New congestion control schemes over wireless networks: Stability analysis," *Proceedings of the* 16^{th} *IFAC World Congress, Prague*, 2005.
[13] A. Abate, M. Chen, and S. Sastry, "New congestion control schemes over wireless networks: Delay sensitivity analysis and simulations," *Proceedings of the* 16^{th} *IFAC World Congress, Prague*, 2005.
[14] S. H. Low and D. E. Lapsley, "Optimzation flow control i: Basic algorithm and convergenece," *IEEE/ACM Trans. on Networking*, pp. 861–875, 1999.
[15] S. H. Low, "A duality model of tcp and queue management algorithms," *IEEE/ACM Trans. on Networking*, Aug 2003.
[16] S. Sastry, *Nonlinear Systems: Analysis, Stability and Control*. Springer Verlag, 1999.
[17] H. Khalil, *Nonlinear Systems (3rd edition)*. Prentice Hall, 2001.
[18] S. Kunniyur and R. Srikant, "A time scale decomposition approach to adaptive ecn marking," *Proceeding of IEEE Infocom*, vol. 1330-1339, Mar 2001.
[19] R. Horn and C. Johnson, *Matrix Analysis*. Cambridge University Press, 1985.
[20] J. Marsden and M. Hoffman, *Elementary Classical Analysis. Second edition*. W.H.Freeman and Company, New York, 1993.

APPENDIX

A. Proof of Lemma 1

Proof: Focusing on (15), the logic of proving the desired result is to show that one $x(t)$ results in one $w(t)$, and conversely one $w(t)$ results in one $x(t)$.

- It is easy to see from (15) that one $x(t)$ maps to a unique $w(t)$.
- Now we show that, given $w(t)$, there is only one set values of $x(t)$ satisfying (15). Given $w(t) = w$, (15) is the maximum for the following strictly concave function of x over $\mathbb{R}^N$

$$U(x) = \sum_{r\in R} w_r \log x_r - \sum_{j\in J} \int_0^{\sum_{s:j\in s} x_s} \left(\epsilon_j + \frac{(y-C_j)^+}{y}\right) dy.$$

The strict concavity implies that the maximum exists over $\mathbb{R}^N$ and is unique; hence there can only be one set values of $x(t) = x$ satisfying (15). Therefore, one $w(t)$ maps to only one $x(t)$.

The relation between $\dot{w}$ and $\dot{x}$ is derived as in [11]. For the last claim of the proposition, refer to known results from [19]. ■

B. Proof of Lemma 2

Proof: Around the equilibrium of the reduced system, let $x_r(t) = x_r^o + z_r(t)$, denote $D(x^o)$ as $\tilde{D}$; after linearization, we have that, $\forall r \in R$,

$$\begin{aligned}\dot{z}(t) &= c\tilde{D}^{-1}\Bigg[\frac{1}{x_r^o} f_r(x^o)\sum_{j\in r}\left(\epsilon_j + g_j(y_j^o)\right) z_r(t) \\ &\quad + \sum_{j\in r}\left(\epsilon_j + g_j(y_j^o)\right)\mu_r^o \sum_{j\in r}\beta g_j'(y_j^o)\sum_{s:j\in s} z_s(t) \\ &\quad + f_r(x^o)\sum_{j\in r} g_j'(y_j^o)\sum_{s:j\in s} z_s(t)\Bigg], \\ &= -c\tilde{D}^{-1}\Bigg[diag\left(f_r(x^o)\right)\tilde{D} + diag\Big(\sum_{j\in r}\left(\epsilon_j + g_j(y_j^o)\right)\Big) \\ &\quad \cdot diag\left(\beta\mu_r^o\right) A^T diag(g_j'(y_j^o)) A\Bigg] z(t), \qquad (17)\end{aligned}$$

where

$$\mu_r^o = \frac{e^{\sum_{j\in r}\ln\left(1+e^{\beta\frac{y_j^o - C_j}{y_j}}\right)}}{\left(1 + e^{\sum_{j\in r}\ln\left(1+e^{\beta\frac{y_j^o - C_j}{y_j}}\right)}\right)^2} > 0, \quad r \in R,$$

and

$$g_j'(y_j^o) = \frac{C_j}{(y_j^o)^2}\frac{e^{\beta\frac{y_j^o - C_j}{y_j^o}}}{1 + e^{\beta\frac{y_j^o - C_j}{y_j^o}}} > 0, \quad j \in J.$$

Denote $E = diag\left(f_r(x^o)\right)\tilde{D} + diag\left(\sum_{j\in r}\left(\epsilon_j + g_j(y_j^o)\right)\beta\mu_r^o\right) A^T diag(g_j'(y_j^o))A$. Then by simple arguments, the system in (17) is stable if and only if $\tilde{D}^{-1}E$ has all positive eigenvalues. We now show that this requirement is verified.

First note that this is equivalent to show $E\tilde{D}^{-1}$ has all eigenvalues be positive since $E\tilde{D}^{-1}$ is similar to $\tilde{D}^{-1}E$. Define $G = diag(\sum_{j\in r}\frac{\epsilon_j + g_j(y_j^o)}{x_r^o})$, then

$$E\tilde{D}^{-1} = diag(f_r(x^o)) + G\cdot diag\left(x_r^o\beta\mu_r^o\right) A^T diag(g_j'(y_j^o))A\tilde{D}^{-1}.$$

At the same time we notice that

$$\begin{aligned}&\tilde{D}\left[A^T diag(g_j'(y_j^o))A\right]^{-1} \\ =\ & \left[G + A^T diag(g_j'(y_j^o))A\right]\left[A^T diag(g_j'(y_j^o))A\right]^{-1} \\ =\ & G\left[\left[A^T diag(g_j'(y_j^o))A\right]^{-1} + \left[diag(\sum_{j\in r}\frac{\epsilon_j + g_j(y_j^o)}{x_r^o})\right]^{-1}\right].\end{aligned}$$

Hence, define the terms inside the brackets as B; we can then have the following expression for $E\tilde{D}^{-1}$:

$$\begin{aligned}E\tilde{D}^{-1} &= diag(f_r(x^o)) + G\cdot diag\left(\beta\mu_r^o\right) B^{-1}G^{-1} \\ &= G\cdot diag(\beta x_r^o\mu_r^o)\left\{diag(\frac{f_r(x^o)}{\beta x_r^o\mu_r^o}) + B^{-1}\right\} G^{-1}.\end{aligned}$$

Now we claim $E\tilde{D}^{-1}$ has all eigenvalues be positive, due to the following three facts:

- $B \succ 0$ since it is a sum of two positive definite matrices. Hence $diag(\frac{f_r(x^o)}{x_r^o\beta\mu_r^o}) + B^{-1} \succ 0$ by the same argument.
- $diag(\beta x_r^o\mu_r^o)\left\{diag(\frac{f_r(x^o)}{x_r^o\beta\mu_r^o}) + B^{-1}\right\}$ has all its eigenvalues to be positive, because it is the product of two positive definite matrices [19];

- $E\tilde{D}^{-1}$ has all eigenvalues be positive, because it is similar to $diag(\beta x_r^o \mu_r^o)\left\{diag(\frac{f_r(x^o)}{\beta\mu_r^o}) + B^{-1}\right\}$.

Eventually, $\tilde{D}^{-1}E$ has all eigenvalues be positive and hence the system in (17) is exponentially stable for arbitrary $\beta > 0$. ■

C. Proof of Theorem 2

Proof: First, any equilibrium (x^*, w^*) of the system in (12) must lie on the equilibrium manifold defined by (15). We also know on this one-to-one mapping manifold, the entire system collapses to a lower dimension reduced system shown in (14). Therefore, it is equivalent to investigate the reduced system for the existence and uniqueness of equilibrium.

Here, we apply the Poincare-Hopf Index Theorem to claim there are at least one equilibrium existed in the reduced system, then apply together Lemma 2 to conclude the number of equilibriums must be one.

Fact 1: (*Poincare-Hopf Index Theorem*) Let $\mathcal{D}$ be an open subset of $\mathbb{R}$ and $\nu : \mathcal{D}^N \to \mathbb{R}^N$ be a smooth vector field, with nonsingular Jacobian matrix $\partial\nu/\partial p$ at every equilibrium p. If there is a $\mathcal{G} \subseteq \mathcal{D}^N$ such that every trajectory moves inward of region $\mathcal{G}$, then the sum of the indices of the equilibria in $\mathcal{G}$ is $(-1)^N$.

To apply Poincare-Hopf Index Theorem, we need to construct a proper vector field and the corresponding region $\mathcal{G}$. For the reduced system, it is equivalent to investigate either $w(t)$ or $x(t)$ as they are connected through an one-to-one mapping.

We claim that the vector field defined by

$$\nu(w(t)) := \dot{w}(t) = c\left([w_r^o, r \in R] - [w_r(t) f_r(x(t)), r \in R]\right) \quad (18)$$

is the one we want. To see that, first note $\nu(w(t))$ can be expressed as a function of $x(t)$, the Jacobian matrix can be expressed as

$$\partial\nu/\partial w = \partial\nu/\partial x \cdot \partial x/\partial w.$$

We have shown that if x^* is an equilibrium of system in (14), then x^* is locally stable, indicating $\partial\nu/\partial x$ is nonsingular at the equilibrium. Also note x and w are one-to-one mapping, hence $\partial x/\partial w$ is nonsingular. Hence $\partial\nu/\partial w$ is nonsingular at the equilibrium.[6]

We now start to construct the necessary region G. First note the following facts:

- if route r is not congested, $g_j(y_j(t)) \le \frac{1}{\beta}\ln 2$; so

$$x_r(t) \ge \frac{w_r(t)}{\sum_{j\in r}(\epsilon_j + \frac{1}{\beta}\ln 2)}.$$

  As we increase $w_r(t)$, $x_r(t)$ will eventually hint $\min_{j\in r} C_j$ and route r is congested (the existence of cross traffic can only help to make the route congested). Hence we claim if $w_r(t)$ is sufficiently large, the route will be congested, regardless of the traffic pattern in the network.

- if route r is congested, at least at one link j on the route, the aggregate arriving rate $y_j(t)$ must exceed the link capacity C_j, therefore

$$f_r(x(t)) \ge 2\frac{e^{\ln 2}}{1 + e^{\ln 2}} - 1 = 1/3.$$

  Hence $w_r^o - w_r(t) f_r(x(t)) < w_r^o - w_r(t)/3$ as long as route r is congested.

Therefore, as $w_r(t)$ becomes sufficiently large, the route r must be congested. There must exist one w_r^{max} such that $w_r^o - w_r^{max} f_r(x) < w_r^o - w_r^{max}/3 < 0$. Then the region $\mathcal{G}$ can be defined as

$$\mathcal{G} = [0, w_1^{max}] \times [0, w_2^{max}] \cdots [0, w_N^{max}].$$

On the boundary of $\mathcal{G}$, we check the flow of the vector field:

- if $w_r(t) = 0$, then easy to see $\dot{w}_r(t) > 0$ according to (18).
- if $w_r(t) = w_r^{max}$, then by the definition of w_r^{max} and according to (18), $\dot{w}_r(t) < 0$.

Therefore, every point on the boundary $\mathcal{G}$ will move inward.

Before we use the Poincare-Hopf Index Theorem, the following Lemma says there are only finite number of equilibria inside $\mathcal{G}$.

Lemma 3: Let M denote the number of equilibriums inside $\mathcal{G}$, and $0 < w_i^e < w^{max}$ represents the ith equilibrium, then $M < \infty$.

Proof: This is because any equilibrium is locally exponentially stable by Lemma 2, and hence is locally unique in an open set around it. The set of equilibriums, denoted as $\mathcal{E} = \{w^e | w_r^o - w_r^e f_r(w^e) = 0, r \in R\}$, is closed and bounded (i.e. compact) since the $w_r^e f_r(w^e)$ is continuous and w^e is bounded. The union of those disjoint open sets, each including one locally unique equilibrium $w^e \in \mathcal{E}$, forms a cover of $\mathcal{E}$. By [20], we claim the number of these disjoint open sets must be finite. Therefore M is finite. ■

Hence by Poincare-Hopf Index Theorem, and notice M is finite, we have the following equations, indicating that there are at least one equilibrium inside region $\mathcal{G}$ and

$$Index(\mathcal{G}) = (-1)^N = \sum_{i=1}^{M} Index(w_i^e),$$

where N is the dimension of $w(t)$.

But every w_i^e is locally stable, hence the Jocobian matrix at the equilibrium w_i^e, denoted by $J(w_i^e)$, has all its eigenvalues be negative. Therefore

$$Index(w_i^e) = sgn(Det(J(w_i^e))) = (-1)^N.$$

Therefore, we can see these two equations imply $M = 1$. Together with the fact that any point outside $\mathcal{G}$ can not be an equilibrium, we conclude there is only one equilibrium for system in (14).

Finally, as the reduced order system has only one unique equilibrium on the equilibrium manifold, we conclude the system (12) has a unique equilibrium, for arbitrary $\beta > 0$. ■

[6] $rank(A) + rank(B) - k \le rank(AB) \le \min(rank(A), rank(B))$.